Little eScience, Big eScience

Panel proposal abstract for the 2009 iConference

Kevin Crowston
Syracuse University School of Information Studies
crowston@ischool.su

David Ribes
Georgetown University
dr273@georgetown.edu

Steve Sawyer
Syracuse University School of Information Studies
ssawyer@syr.edu

Andrea Wiggins
Syracuse University School of Information Studies
awiggins@syr.edu

ABSTRACT

eScience refers to the use of information and communications technologies (ICT) to support scientific work. eScience is relevant to the iConference as eScience is both a potential mode of research and a topic of research that is particularly relevant to iSchools. However, we note that the most visible applications of eScience are in the "big" sciences and it is not clear if or how these ideas can be transferred to the "little sciences". We propose a panel to address this question. The panelists will discuss how eScience ideas might apply in the "little" sciences, and some of the challenges and open research questions involved in such an application.

1. WHAT IS ESCIENCE

eScience refers to the use of information and communications technologies (ICT) to support scientific work. Nentiwich defines eScience as "scholarly and scientific research activities in the virtual space generated by the networked computers and by advanced information and communication technologies" [2, p. 22]. eScience is also known as eResearch, cyberscience or even Science 2.0, and the supporting systems as cyber-infrastructure or scientific collaboratories.

Because there are many types of ICT and many kinds of scientific activity to which they can be applied, eScience encompasses a broad range of distinct cyber-infrastructure applications. For many, cyber-infrastructure means high performance computing, e.g., grid computing to support analyses of large volumes of data as well as simulations, which have become for many the third approach to science, between theory and experiment. But cyber-infrastructure also includes Internet-enabled applications to connect scientists to a variety of resources: data, knowledge and other researchers. Data might come from instruments directly connected to the Internet, a shared instrument, or from structured data repositories in a community data system. Knowledge might be captured in digital libraries of journal articles or, increasingly, of article preprints. Connections between data and knowledge can also be made explicit; a simple step is to link publications to data sources and vice versa. More interestingly, scientific workflow tools can be used to capture analysis steps explicitly in an executable format, enabling reuse and sharing of analyses. Increasingly, eScience applications are built using semantic web technologies that enable automated reasoning about scientific knowledge. Various fields have developed ontologies (structured or controlled vocabularies) that describe the objects of study, such as the human anatomy or types of cells in medicine and the relations among them.

Finally, because science is not a solitary pursuit, eScience applications can also include groupware to support scientific collaboration. These applications can range from simple email and mailing lists connecting collaborators, to digital libraries of various scopes, to newer collaborative applications such as wikis, shared document editors or semantic web reasoners. These tools can support virtual collaborations of different scales and degrees of formality. For example, discussion boards linked to papers in preprint archives have been proposed to augment or even replace some of the functions of conventional peer review [2]. To the above list of technologies, we must add the necessary social arrangements needed to make the technology successful.

2. "BIG" AND "LITTLE" ESCIENCE

In order to understand the implications of eScience, it is important to delineate the kinds of research for which eScience approaches are likely to be feasible and useful. A distinction that we explore in this panel is between "big science" and "little science" [1]. "Big science" refers to scientific projects that draw on multiple disciplines to address a broad set of goals, which are often set by a committee that then selects the researchers to carry out the work. Big science increasingly demands eScience methods; the cost of creating knowledge has increased dramatically for many scientific ventures, with projects requiring shared infrastructure such as supercolliders, telescopes, research ships and other large instruments, plus the engineers and administrators needed to keep a big science project running. As a result, more researchers are dependent on the outputs of the scientific infrastructure.

The fundamental nature of the data produced in big science has also changed; petabytes of data and teraflops of processing are now normal operating conditions for many natural sciences. Growing arrays of sensors and instrumentation produce more data streams to fuse for analysis, creating additional challenges for researchers who now have to merge multiple high-volume data products. In addition, funders are often requiring that the data generated by large research efforts to be shared with the research community more widely, making eScience contributions by "outsiders" an increasingly likely source of new knowledge discovery. Because big science has been working on adapting to eScience technologies for some time, many fields in the natural sciences have developed standards to enable efficient use of the data and technology. While learning these standards is a barrier to entry for newcomers into the field, it may prove a lower barrier than long-term laboratory apprenticeship, and the standards themselves offer a more complete, carefully vetted documentation of the state of scientific practice than most individual research experiences can encompass.

By contrast, "little science" refers to a single investigator working on projects of their own choosing with relatively modest support, such as a graduate student or two. In little science, the advantages of eScience methods are less clear. Little science is less likely to have standardized work in a way that allows smooth re-use of data and analysis tools. While in principle the same technologies and practices have the same sort of potential for enabling discovery in little science, the reality is that in small research communities, data and technologies have generally been created for the individual research groups' goals, and the infrastructure that enables transferability and interoperability of data and analysis technologies may not yet be available. Nonetheless, it is worth noting that even within little science research groups, standardization and documentation is valuable, particularly as students graduate or as collaboration opportunities arise. It seems likely that little science will require more time to adopt the social infrastructure of practice that big science has already developed of necessity, more coaxing to achieve community buy-in of the principles and practices in the absence of funding-driven collaboration of eScience, and more flexibility built into the standards as they evolve to allow for the greater variability of scientific investigation that is accomplished through use of qualitative data, for example. At the same time, little science has the advantage of the examples of infrastructure developed by big science. Strategic efforts by leaders in specialized fields to adopt and adapt the existing practices and technologies has the potential to produce dramatic results.

3. PLAN FOR PANEL

The proposed panel will start with an overview of eScience and of [1]'s distinction between big and little science. Panelists will then discuss issues that face the little sciences in adopting eScience ideas, discussing in turn access to data, research products and collaborators. Each presentation will consider both how such issues affect our own use of eScience as well as possible research questions that iSchools might address. Ample time will be left for audience questions and contributions.

4. ACKNOWLEDGMENTS

This research was partially supported by US NSF Grants 05-27457 and 07-08437.

5. REFERENCES

[1] D. J. de Solla Price. *Little Science, Big Science.* Columbia University, New York, 1963.

[2] M. Nentwich. *Cyberscience: Research in the Age of the Internet.* Austrian Academy of Sciences, Vienna, 2003.